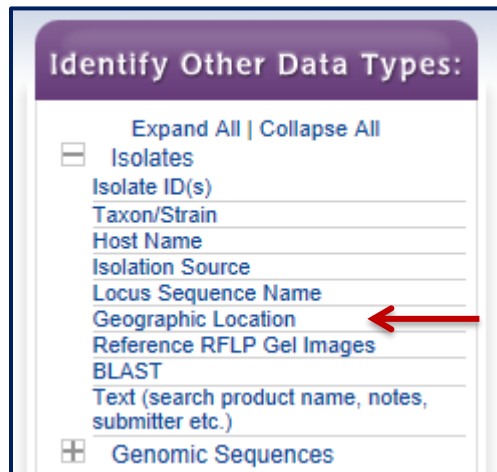
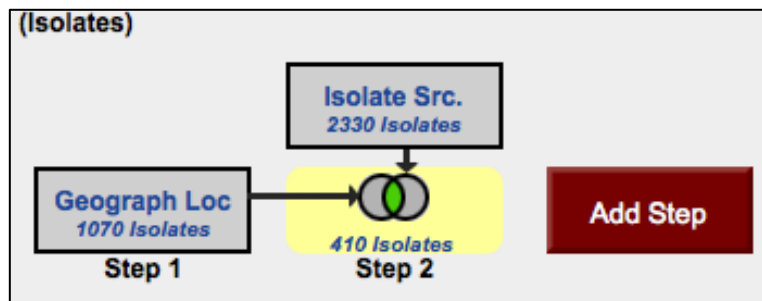


Exploring Isolate Data

1. Exploring isolates in *Cryptosporidium* and using the alignment tool. (<http://www.cryptodb.org>)
 - a. Identify all *Cryptosporidium* isolates from Europe. (hint: search for isolates by geographic location in the “Identify Other Data Types” section).



- b. How many of the *Cryptosporidium* isolates collected in Europe were isolated from feces? (hint: add another isolate search step - isolation source).

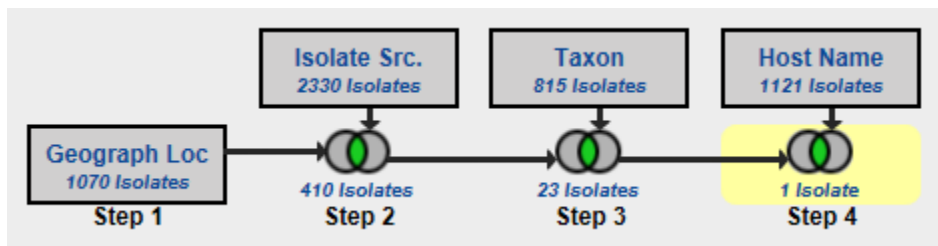


- c. What is the general distribution of these isolates in Europe? (hint: you can do this quickly in two ways: sort the geographic location column by clicking on the sort arrows, then look at the represented countries; or use the histogram tool on the Geographic Location column; or use the “Isolate Geographical Location” tab to view a map and results summary table).

Sort by clicking on the arrows, or view data in histogram tool

Country	Number of Isolates	Isolate Type	Latitude	Longitude
Belgium	1	Sequencing Typed	50.503887	4.469936
Czech Republic	73	Sequencing Typed	49.817492	15.472962
Germany	90	Sequencing Typed	51.165691	10.451526
Ireland	4	Sequencing Typed	53.41291	-8.24389
Italy	11	Sequencing Typed	41.87194	12.56738
Lithuania	1	Sequencing Typed	55.169438	23.881275
Netherlands	41	Sequencing Typed	52.132633	5.291266

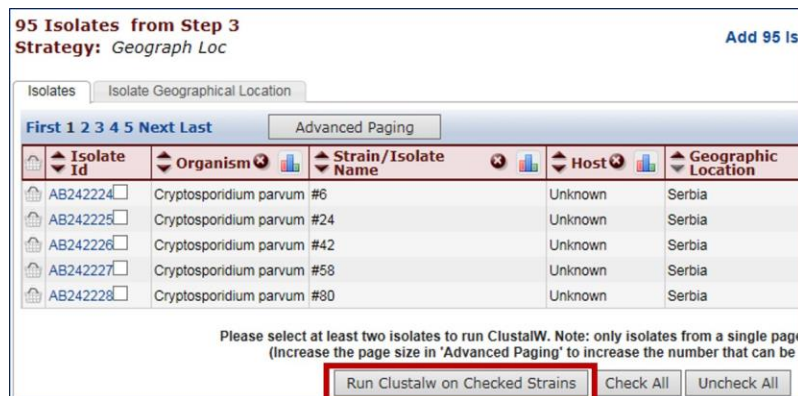
- d. Out of those in step ‘b’, how many are unclassified *Cryptosporidium* species? (hint: add another isolate search step and select taxon/strain then select the unclassified isolates)
- e. How many of step ‘d’ isolates originated from humans?



- f. How many of the isolates in step 'b' were typed using GP40/15 (GP60)? (hint: you can insert a step within a strategy. Click "edit" on the step of interest then select "Insert step before").



- g. Compare some of these isolates using the multiple sequence alignment tool (ClustalW). Do you see any sequences with insertions or deletions?



- h. Take a look at the 'guide tree' that was built to help generate this alignment. The guide tree is located below the ".dnd" text located at the end of your multiple sequence alignment file. It may look something like the text below. The dendrogram is in a "newick" file format.

```
(
AB242228:0.00305,
(
AB242229:-0.00778,
(
(
AY508961:0.86194,
EF576957:-0.01467)
:0.03332,
EF576958:0.02143)
:0.03432)
:0.00778,
EF576956:0.00000);
```

Note: the beginning “ (“ and closing “ ; ” are important parts of the file format. You can use your mouse to select the text at the end of your file, copy it, and paste it into the box at the [tree viewer site](#) (remove the sample file in the box before adding your own sequence). Click on “view tree” to visualize the tree encoded in the text.

The screenshot shows the 'Newick Viewer' web application. On the left is a 'Main Menu' with various options like 'Tree viewer', 'Tree builder', and 'Tree inference'. The main content area has a title 'Newick Viewer' and a description: 'Newick Viewer allows you to visualize a tree coded by its Newick string. Hierarchical, Axial and Radial types of tree drawing are available.' Below this is a text input box containing a Newick string:

```
( AB242228:0.00305, ( AB242229:-0.00778, ( AY508961:0.86194, EF576957:-0.01467) :0.03332, EF576958:0.02143) :0.03432) :0.00778, EF576956:0.00000);
```

 Below the input box are radio buttons for 'Sequences file', 'Pasted', and 'Choose File', with 'Pasted' selected. At the bottom, there are buttons for 'View Tree', 'Reset', and 'Clear'. The 'View Tree' button is highlighted with a red box.

Change the isolates that you selected for alignment - how does the tree change?
Do isolates from the same country cluster together?

2. Typing an unclassified *Cryptosporidium* isolate. (<http://www.cryptodb.org>)

- a. You have just finished sequencing part of the 18S small subunit ribosomal RNA gene from isolates you retrieved from a *Cryptosporidium* outbreak at a public swimming pool in Uppsala. The sequence was identical from all the isolates and is pasted below. Can you use CryptoDB to get an idea of which reference isolate this is most similar to? (hint: blast your sequence against the reference isolates).

```
AAGCTCGTAGTTGGATTTCTGTTAATAATTTATATAAAATATTTTGATGAATATTTATAT
AATATTAACATAATTCATATTACTATATATTTTAGTATATGAAATTTTACTTTGAGAAAA
TTAGAGTGCTTAAAGCAGGCATATGCCTTGAATACTCCAGCATGGAATAATATTAAGAT
TTTTATCTTTCTTATTGGTTCTAAGATAAGAATAATGATTAATAGGGACAGTTGGGGGCA
TTTGTATTTAACAGTCAGAGGTGAAATTCCTAGATTTGTTAAAGACAACTAATGCGAAA
GCATTTGCCAAGGATGTTTTTCATTAATCAAGAACGAAAGTTAGGGGATCGAAGACGATCA
GATACCGTCGTAGTCTTAACCATAAACTATGCCAACTAGAGATTGGAGGTTGTTCCCTTAC
TCCTTCAGCACCTTA
```

- b. You can get to the isolate BLAST page from the home page (BLAST link under popset isolate sequences) select “BLAST”. Configure the BLAST search page: select isolates and make sure only the reference isolates are selected in the target organism window.
- c. Paste the DNA sequence in the input window and select the blastn program. Click on “Get Answer”.

The screenshot shows the CryptoDB BLAST search interface. Red arrows point to the following fields:

- Target Data Type:** Radio buttons for Transcripts, Proteins, Genome, EST, ORF, and **Isolates** (selected).
- BLAST Program:** Radio buttons for **blastn** (selected), blastp, blastx, tblastn, and tblastx.
- Target Organism:** A list of taxonomic groups with checkboxes. **Cryptosporidiidae SSU_18srRNA Reference Isolates** is checked. Other groups include Chromerida, Cryptosporidiidae, and Gregarinidae.
- Input Sequence:** A text area containing the DNA sequence from the previous block.

Below the input sequence, there are fields for:

- Expectation value: 10
- Maximum descriptions/alignments (V-B): 50
- Low complexity filter: no

At the bottom, there is an "Advanced Parameters" section with a "Get Answer" button and a text input field for naming the search.

- d. Explore your results. Based on the similarity which reference isolate is this one closest to?

```

AF093490 | organism=Cryptosporidium_parvum | description=Cryptos... 862 0.0
AF093491 | organism=Cryptosporidium_parvum | description=Cryptos... 817 0.0
AF112571 | organism=Cryptosporidium_parvum | description=Cryptos... 813 0.0
AF112572 | organism=Cryptosporidium_parvum | description=Cryptos... 809 0.0
AF115378 | organism=Cryptosporidium_wrairi | description=Cryptos... 809 0.0
AF159110 | organism=Cryptosporidium_parvum | description=Cryptos... 801 0.0
AF112574 | organism=Cryptosporidium_meleagridis | description=Cr... 799 0.0
AY737573 | organism=Cryptosporidium_environmental_sequence | des... 781 0.0
AY737560 | organism=Cryptosporidium_environmental_sequence | des... 761 0.0
AF262330 | organism=Cryptosporidium_sp. | description=Cryptospor... 753 0.0

>AF093490 | organism=Cryptosporidium_parvum |
description=Cryptosporidium parvum strain Bovine C.
parvum genotype (BOH6) small subunit ribosomal RNA gene,
complete sequence. | length=1746
Length = 1746

Score = 862 bits (435), Expect = 0.0
Identities = 435/435 (100%)
Strand = Plus / Plus

Query: 1 aagctcgtagttggatttctgttaataatttatataaaaatattttgatgaatatttatat 60
|||
Sbjct: 601 aagctcgtagttggatttctgttaataatttatataaaaatattttgatgaatatttatat 660

Query: 61 aatattaacataattcatattactatatatatttttagtatatgaaattttactttgagaaaa 120
|||
Sbjct: 661 aatattaacataattcatattactatatatatttttagtatatgaaattttactttgagaaaa 720

Query: 121 ttagagtgccttaagcagccatagccttgaatactccagcatggaataatattaaagat 180
|||
Sbjct: 721 ttagagtgccttaagcagccatagccttgaatactccagcatggaataatattaaagat 780

Query: 181 ttttatctttcttattgggttctaagataagaataatgattaatagggacagttgggggca 240
|||
Sbjct: 781 ttttatctttcttattgggttctaagataagaataatgattaatagggacagttgggggca 840

Query: 241 tttgtatttaacagtcagaggtgaaattcttagatttggttaagacaaaactaatgcgaaa 300
|||
Sbjct: 841 tttgtatttaacagtcagaggtgaaattcttagatttggttaagacaaaactaatgcgaaa 900

Query: 301 gcatttgccaaggatgttttcattaatcaagaacgaaagttaggggatcgaagacgatca 360
|||
Sbjct: 901 gcatttgccaaggatgttttcattaatcaagaacgaaagttaggggatcgaagacgatca 960

```